The Annals of Statistics
2000, Vol. 28, No. 3, 713–733

TRACKING A SMOOTH FAULT LINE IN A RESPONSE SURFACE

BY PETER HALL AND CHRISTIAN RAU

Australian National University and CSIRO and Australian National University

We suggest a sequential, or "tracking," algorithm for estimating a smooth fault line in a response surface. The method starts with an approximation to a point on the line, and from there the line is tracked as it meanders through the plane. The technique differs from recent approaches in that it does not require a large part of the plane to be searched for evidence of a fault line. This offers potential computational savings, and produces a method that is invariant under rotations of coordinate axes (except insofar as a rotation might affect the estimated starting point, and the relative orientation of the grid on which calculations are done). That feature is important if design points are not located on a regular grid. We investigate properties of the method under very general conditions on the design, allowing Poisson cluster processes, jiggled grid processes and deterministic, regular lattices. Uniform rates of convergence are derived in all these settings, for the case of noisy data, and shown to be within logarithmic factors of optimal pointwise convergence rates in the no-noise setting.

1. Introduction. The topic of estimating a fault line in a regression problem with bivariate design has received considerable recent attention. For example, in the context of square lattices a maximum-likelihood approach has been adopted by Rudemo and Stryhn (1994), a pointwise procedure has been suggested by Qiu and Yandell (1997), and a multistage method has been proposed by Qiu (1998). Special features of the case where design variables are on a regular lattice have been addressed by Hall and Raimondo (1997a, b, 1998).

Recently introduced techniques, such as those of Qiu and Yandell (1997) and Qiu (1998), involve computing diagnostics in places that can be some distance from the fault line and then combining the conclusions of different diagnostic analyses so as to produce an estimator. In the present paper we develop an alternative, "tracking" method. It is designed to follow, or track, a fault line once the algorithm has been given a point that lies near the line. With high probability the new method produces an estimator which never strays far from the fault line.

An advantage of this approach, apart from potential savings in computation, is that it is able to follow a wiggly fault line as it meanders almost arbitrarily across the plane. Methods that first make a global assessment of potential fault lines in the data, and then try to piece this information together

Received August 1998; revised August 1999.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62H05.

Key words and phrases. Boundary estimation, change point, edge detection, frontier analysis, image analysis, jump, kernel methods, least squares, smoothing.

to produce a single curve, can become confused when a fault line doubles back towards itself. Moreover, some contemporary methods are applicable only to square-lattice designs, and would require major modification if they were to be used for irregularly distributed designs. Additionally, methods or theoretical arguments that require a fault line to admit a functional equation, for example of the form $y = g(x)$, have to be applied in a piece-by-piece way in regions where the equation would be multiple valued. In addition to not suffering from these problems, a tracking approach produces results that, modulo choice of starting point and of the grid on which computations are done, are independent of choice of coordinate axis. This is particularly relevant when the design points are not on a lattice. A case in point is estimation of benthic impacts (e.g., sediment particle sizes, concentrations of hydrocarbons and selected metals and biological information) from undersea drill cuttings. The drill sites are not located at vertices of a regular lattice, and concentrations, etc. can exhibit fault lines due to sharp changes in geological features on the ocean floor.

In addition to introducing and trying out a tracking method and describing its properties, one of our aims in this paper is to bring out differences between rates of approximation that are obtainable for different types of point processes. Hall and Raimondo (1997a) showed that, even for a problem as simple as estimating a straight-line fault across a square lattice, and even when there is no noise and the image is only black and white, the rate of approximation depends intimately on the nature of the line. If its slope is rational relative to the lattice axes, if we observe the line within a fixed region of the plane and if the lattice in that region contains roughly ν points per unit area, then, unless the line actually passes through one or more lattice points, the optimal rate at which its location can be estimated is $O(\nu^{-1/2})$ as $\nu \to \infty$. However, if the line has irrational gradient then, depending on the "type" of the irrational number (e.g., whether it is a quadratic irrational), the line can be estimated with error as low as $O(\nu^{-1})$.

The latter rates are similar to those obtained for a straight-line fault when design points come from a Poisson process with intensity ν. However, for both lattice designs and Poisson-distributed designs the optimal rate is slower when the fault line is curved. For example, minimax-optimal rates in the case of Poisson-distributed points, when there is no noise and the fault line has bounded curvature, are known to be $O(\nu^{-2/3})$, with a logarithmic factor in the case of uniform rates; see Korostelev and Tsybakov (1993), particularly Section 5.3 and Theorem 5.3.3, and Mammen and Tsybakov (1995).

As Mammen and Tsybakov note, the estimators that have been shown to achieve these rates are not really practical, for a variety of reasons. In the case of noisy data the tracking-type estimators proposed in the present paper are at once practicable and able to achieve optimal rates up to at most a logarithmic factor. The problem of fault-line estimation is closely related to that of edge detection in image analysis [e.g., Marr and Hildreth (1980), Huertas and Medioni (1986)]. The connection between techniques for image analysis and statistical methods based on smoothing has been drawn by, for example,

Titterington (1985a, b) and Cressie [(1993), pages 528–530]. Fault-line estimation has links to boundary-estimation problems in other contexts, for example, in econometrics [e.g., Seiford (1996), Kneip, Park and Simar (1998), Gijbels, Mammen, Park and Simar (1999)]. There is also a strong connection to change-point analysis; see, for example, Müller (1992), Eubank and Speckman (1994) and Müller and Song (1994).

Section 2 describes our tracking method and illustrates its implementation. Theoretical properties are presented in Section 3, with technical arguments deferred to Section 4.

2. Methodology and numerical implementation. Let $\mathscr{C}$ denote a smooth, contiguous, rectifiable curve in the plane, of length $l > 0$ and not intersecting itself. (However, we allow the possibility that the curve is closed.) Suppose the locus of points on $\mathscr{C}$ is determined by the functions $(x(s), y(s))$, where s denotes the distance along $\mathscr{C}$ from a point Q at one end of the curve; and assume that $x(\cdot)$ and $y(\cdot)$ are smooth. No matter whether $\mathscr{C}$ is open or closed we can distinguish left- and right-hand sides of $\mathscr{C}$, with parity determined by tracing $\mathscr{C}$ in the direction of increasing s. By convention, neither the left nor the right side of $\mathscr{C}$ includes any part of $\mathscr{C}$ itself.

Consider a response surface with formula $z = f(x, y)$, where x, y, z are scalars and the function f is smooth except for a fault line of which $\mathscr{C}$ is at least a part. We observe noisy values of the response surface, generated by the model

$$Z_i = f(X_i, Y_i) + \varepsilon_i,$$

where the sequence of pairs (X_i, Y_i) represents a realization of a point process in the plane, and conditional on these points, the errors ε_i are independent and have zero mean. We employ these data to construct a tracking estimator $\widehat{\mathscr{C}}$ of $\mathscr{C}$, as follows.

Assume we have a starting point $\widehat{Q}$ and a starting estimate $\hat{\theta}$ of the orientation of the curve at $\widehat{Q}$. It may be supposed that all potential estimates of points on $\mathscr{C}$ are confined to a square grid $\mathscr{G}$, and likewise that estimates of the orientation of $\mathscr{C}$ are restricted to a discrete grid $\mathscr{G}'$. (If the process of points (X_i, Y_i) was on a lattice then generally $\mathscr{G}$ would be much finer than this.) Suppose we have constructed $\widehat{\mathscr{C}}$ as far as a point P with coordinates (x, y), where our estimate of the tangent to $\mathscr{C}$, in the direction of travel along $\mathscr{C}$, is a unit vector ω. We show how to construct the next point, P' say, with coordinates (ξ, η) and tangent estimate in the direction of another unit vector, θ.

Let $\mathscr{U}(\xi, \eta, \theta)$ denote the line through P' in the direction of the unit vector θ, and let $g(u, v|\xi, \eta, a, b, \theta)$ be the function that equals a (respectively, b) to the left (right) of $\mathscr{U}(\xi, \eta, \theta)$. Let $\|\cdot\|$ denote the Euclidean norm, let $\sum_i^{(0)}$ denote summation over all pairs (X_i, Y_i), let K be a smooth, nonnegative, univariate kernel function, and let $h > 0$ be a bandwidth. Choose (ξ, η, θ)

to minimize

$$S(\xi, \eta, \theta) = \inf_{-\infty<a,b<\infty} \sum_i^{(0)} \{Z_i - g(X_i, Y_i|\xi, \eta, a, b, \theta)\}^2 \times K\{\|(x, y) - (X_i, Y_i)\|/h\}, \tag{2.1}$$

among those four or five of the eight grid neighbors (ξ, η) of (x, y) which are such that the unit vector in the direction of the line from (x, y) to (ξ, η) has a nonnegative dot-product with ω. Choose the sign of the associated unit vector θ so that $\omega \cdot \theta \geq 0$, and further limit the choice of θ to those points in $\mathscr{G}'$ that are contained in the arc centered at ω that is $O(h^\varepsilon)$ radians wide, for some $\varepsilon > 0$.

Our results and their proofs may be generalized to the case of arbitrarily fine grids $\mathscr{G}, \mathscr{G}'$, and even to the continuum. The method discussed theoretically in Section 4 is simplified so that technical arguments are relatively transparent, and so that the number of steps needed to traverse the curve is reasonably small without hindering accuracy. Therefore it is applied to relatively coarse grids $\mathscr{G}, \mathscr{G}'$; finer grids produce methods that are closer to being truly rotationally invariant.

High-order methods, for example based on capturing curvature as well as slope in local approximations to $\mathscr{C}$, are also feasible. They require more complex local approximations to the response surface, however, if full benefit is to be gained in terms of convergence rates. For example, if we model curvature of $\mathscr{C}$ then we should also model the tangent planes to the response surface on either side of $\mathscr{C}$, as well as the surface heights. This means that, overall, eight (rather than three) parameters need to be fitted locally. That reduces the attractiveness of the procedure.

To explore numerical performance of the method we investigated several levels of noise degradation of various surfaces defined by response functions and considered designs in the form of either a regular lattice or a homogeneous Poisson process. An explicit formula can be derived for the quantity $S(\xi, \eta, \theta)$ defined at (2.1); see (4.1). To minimize S we used an algorithm based on golden-section search, described for example, in Press, Teukolsky, Vetterling and Flannery, (1992). Writing ω, representing a unit vector, for the previously estimated real-valued tangent direction, the search for a minimum was carried out within an arc approximately $\pi/10$ radians wide, centered at ω.

We experimented with a number of variants of the methodology suggested above. One was to take (ξ, η) equal to (x, y) in (2.1), and move an amount δ in the direction of the fitted value of θ, rather than move directly to a neighboring point of the grid $\mathscr{G}$. (We took $\delta = 0.005$.) Another was to mix this approach with the method suggested at (2.1), thereby refitting location. We found that it was adequate, and saved time, to refit location at points distant δ and $\delta/2$ on either side of (x, y) on the line that passed through (x, y) and was perpendicular to the estimated orientation of the curve at (x, y).

In low-curvature parts of $\mathscr{C}$ the cycle length between successive location refittings could be varied within a wide range, from 1 to about 10 steps, without

having any noticeable effect on performance. However, refitting at *each* step was noticeably beneficial in places of high curvature. A related matter is that of estimating θ, which represents a derivative and so is inherently more prone to error than an estimate of $\mathscr{C}$. Nevertheless, these difficulties could be overcome in part by using a smaller value of δ in places where $\mathscr{C}$ has higher curvature. It might even be appropriate to use cycle length and δ as spatially varying tuning parameters. In principle, curvature could be estimated from values of $\widehat{\mathscr{C}}$ computed in earlier steps, and used to select these quantities at the current location.

For definiteness, the curves shown here were drawn using an algorithm that refitted location at each step. Also, to smooth out sharp turns to some extent we derived $\hat{\theta}$ from a moving average of the present step and two previous ones, with weights 0.6, 0.3 and 0.1.

It was found efficacious to smooth over an ellipse, rather than a disc, with its shorter axis in the direction of motion. We took the two axis lengths to be 0.08 and 0.12.

For brevity and simplicity we present here only our results for a single response surface, illustrated in panel (a) of Figure 1 and given in polar coordinates by

$$f(r, \phi) = I\big[8r \le \exp\{\cos(3\phi)\} + \sin(\phi/2)\big], \qquad 0 \le \phi \le 2\pi,$$

where I denotes the indicator function. This example illustrates how the algorithm copes with "meandering functions" that do not admit simple representations in Cartesian coordinates. We used the biweight kernel, $K(x) = (15/16)(1 - x^2)^2$ for $|x| \le 1$, and took noise to be Normal $N(0, \sigma^2)$, where $\sigma = 0.5$ or 0.75. Design points were either on a regular 330×330 lattice within the square $[-0.55, 0.55] \times [-0.55, 0.55]$, or in the form of a homogeneous Poisson process with the same expected number of points per unit area. Typical realizations of lattice-distributed, noisy data are illustrated in panels (b) and (c) of Figure 1, corresponding to the two respective values of σ.

Our first objective was to estimate $\mathscr{C}$ over its full length. For that purpose, and for definiteness, we chose the single starting point $\widehat{Q} = (0.2, 0.08)$ (which is located close to $\mathscr{C}$) and the approximate tangent direction $\theta(0) = \pi$. If $\widehat{Q}$ is selected by eye, and if the design points form a regular lattice, then one or more of the design points may lie on the line that divides the kernel domain into two half-discs, causing ambiguity. While in theory this causes only negligible problems, in practice the possibility of these ties should be recognised.

Panels (a)–(d) of Figure 2 are to be read in pairs, the left-hand side showing the case of a lattice design and the right-hand side, the case of Poisson-distributed design. Each pair of panels corresponds to a different value of σ. Six typical realizations are superimposed in each panel, and show that performance deteriorates only slightly when $\sigma = 0.75$. From the displays, the conclusion can be drawn that even those parts of $\mathscr{C}$ where absolute curvature briefly exceeds an absolute value of 30, cause few problems. However, we found in numerical work not given here that erratic behavior of the estimates begins

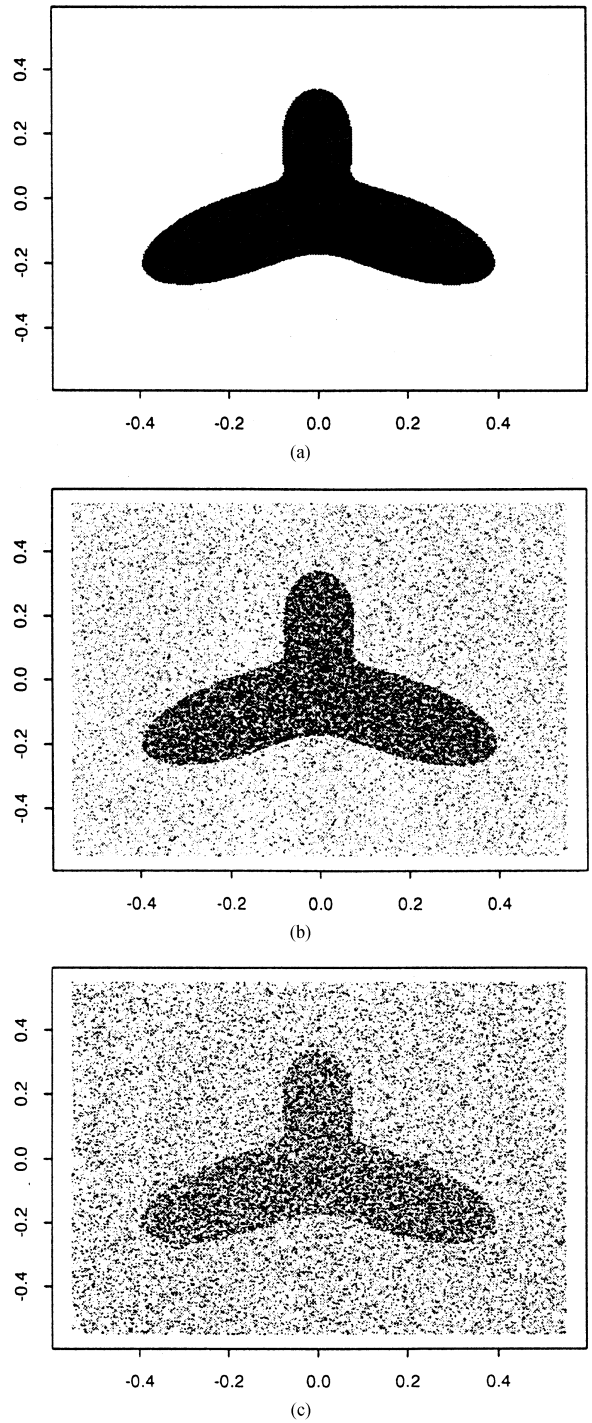

FIG. 1. *Panel* (a) *shows the original response surface, and panels* (b) *and* (c) *depict the superposition of Normal* $N(0, \sigma^2)$ *noise with* $\sigma = 0.5$ *and* $\sigma = 0.75$, *respectively, in the case of lattice design. The lattice edge width is* $1/300$.

to occur regularly for $\sigma = 0.75$ if the previously mentioned auxiliary devices are not employed. Moreover, we found generally that higher noise levels, say $\sigma = 1.0$, most often led to unsatisfying results in the nonstraight parts of $\mathscr{C}$. It would be possible, however, to improve performance here by further increasing the intensity of the point process.

Finally, to conclude this numerical example, we investigated the sensitivity of the algorithm with respect to the starting point $\widehat{Q}$. We continued to use the previous regression surface, focusing now on the higher noise level $\sigma = 0.75$ and the case of Poisson-distributed design. Figure 3 displays our simulation results. Panels (a)–(d) depict behavior of three typical estimates in the respective cases $\widehat{Q} = (q + 0.08\,k, 0)$ with $q = e/8 \approx 0.34$ and $k \in \{-2, \ldots, 1\}$. In this part of our example we smoothed over discs (instead of ellipses) with radius $h = 0.08$, and for illustration we have added a disc of that radius, centered at $\widehat{Q}$, to our displays. Also, on this occasion we tracked the fault line on both sides of the starting point. The results indicate that the algorithm is sensitive to choice of $\widehat{Q}$, although it readily tracks the fault line once the latter has been found. In practice it is a good idea to experiment with different, neighboring starting points.

3. Theoretical properties. We assume that $\mathscr{C}$ is traced out within a compact rectangle $\mathscr{R}$, with its ends at points Q_L and Q_R on the left- and right-hand sides of $\mathscr{R}$, respectively, and that these sides have no other point with $\mathscr{C}$ in common. Let Q denote the first point on $\mathscr{C}$ that is distant h from the left-hand side of $\mathscr{R}$. We start tracing our estimator $\widehat{\mathscr{C}}$ at a point $\widehat{Q}$ distant h from the left-hand side of $\mathscr{R}$, representing an approximation to Q, and stop as soon as we get within h of the other side of $\mathscr{R}$. See Remark 3.4 for methods for calculating both $\widehat{Q}$ and a starting orientation.

The estimator is the piecewise-linear curve defined by joining successive estimates of points on the curve. For simplicity of exposition we take these points to be vertices (ξ, η) of a square lattice $\mathscr{G}$ with edge width $B_1 h^2$, for some $B_1 > 0$, traced out within $\mathscr{R}$. We take the set of candidates for unit vectors θ to be defined by at least $B_2 h^{-1}$ points regularly spaced around the circumference of the unit circle, centered at the origin, where $B_2 > 0$.

Next we specify conditions $(\mathrm{C}_{\mathrm{rs}})$ on the response surface. Let a general point P on the fault line $\mathscr{C}$ be represented by $(x(s), y(s))$, for $0 \le s \le l$ say, where $l < \infty$, Q_L and Q_R are represented by $(x(0), y(0))$ and $(x(l), y(l))$, respectively, and s denotes the distance of P along $\mathscr{C}$ from Q_L. Assume $\mathscr{C}$ does not self-intersect, and that the first two derivatives of $x(\cdot)$ and $y(\cdot)$ are uniformly bounded on $[0, l]$. We assume too that f and each of its first derivatives is uniformly bounded in the intersection $\mathscr{R}_L$ of $\mathscr{R}$ with the left-hand side of $\mathscr{C}$, and also in the intersection $\mathscr{R}_R$ of $\mathscr{R}$ with the right-hand side; and that if, for $(x, y) \in \mathscr{C}$, we define $f_L(x, y)$ to equal the limit of $f(x_1, y_1)$ as (x_1, y_1) converges to (x, y) through values in $\mathscr{R}_L$, and $f_R(x, y)$ to be the limit of $f(x_1, y_1)$

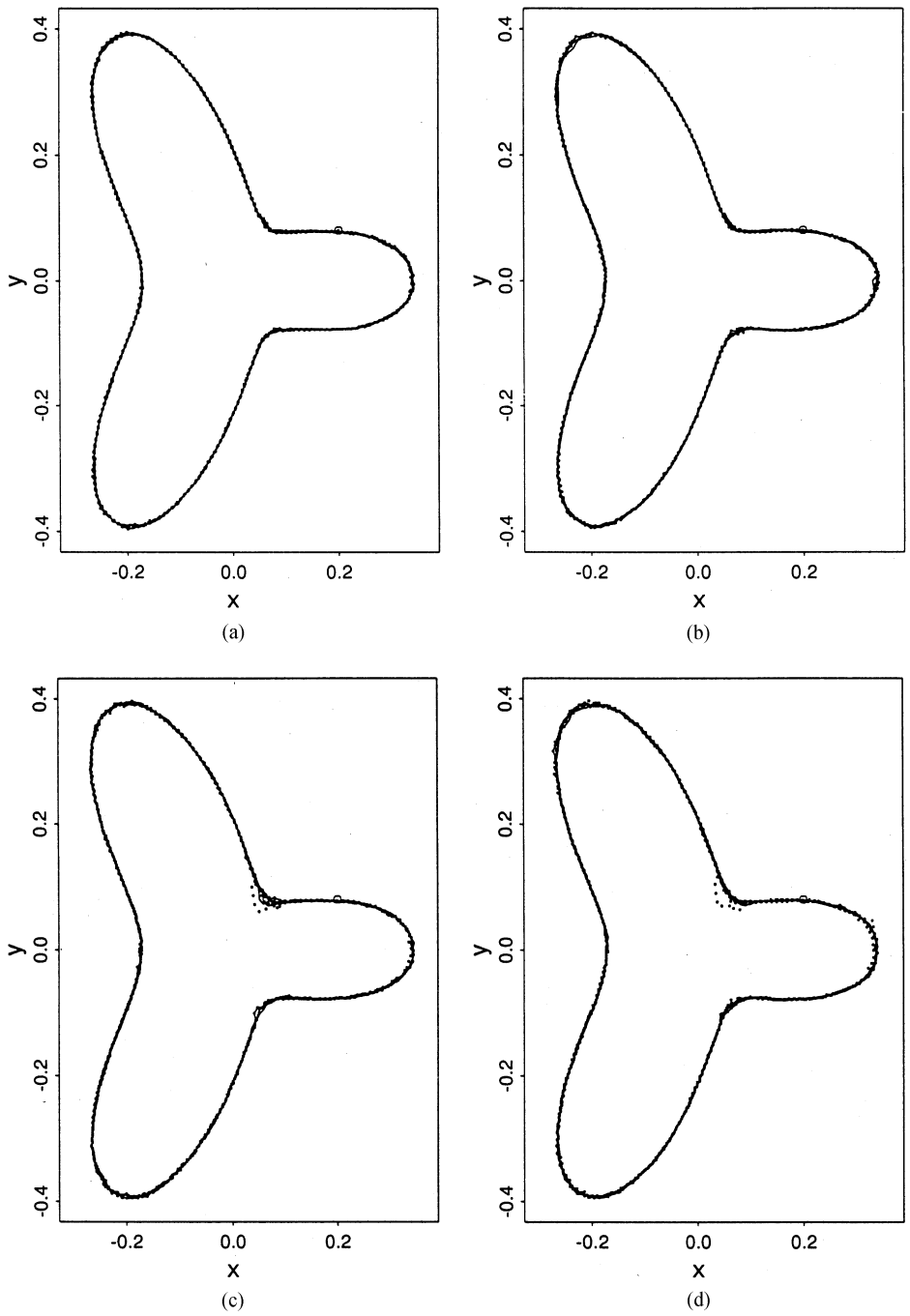

FIG. 2. *Performance of the algorithm is illustrated for* $N(0, \sigma^2)$ *noise, with* $\sigma = 0.5$ *in panels* (a) *and* (b), *and* $\sigma = 0.75$ *in panels* (c) *and* (d). *Panels* (a) *and* (c) *depict the case of gridded design, while panels* (b) *and* (d) *are for the case of Poisson-distributed design. The starting point* $\widehat{Q} = (0.2, 0.08)$ *is encircled.*

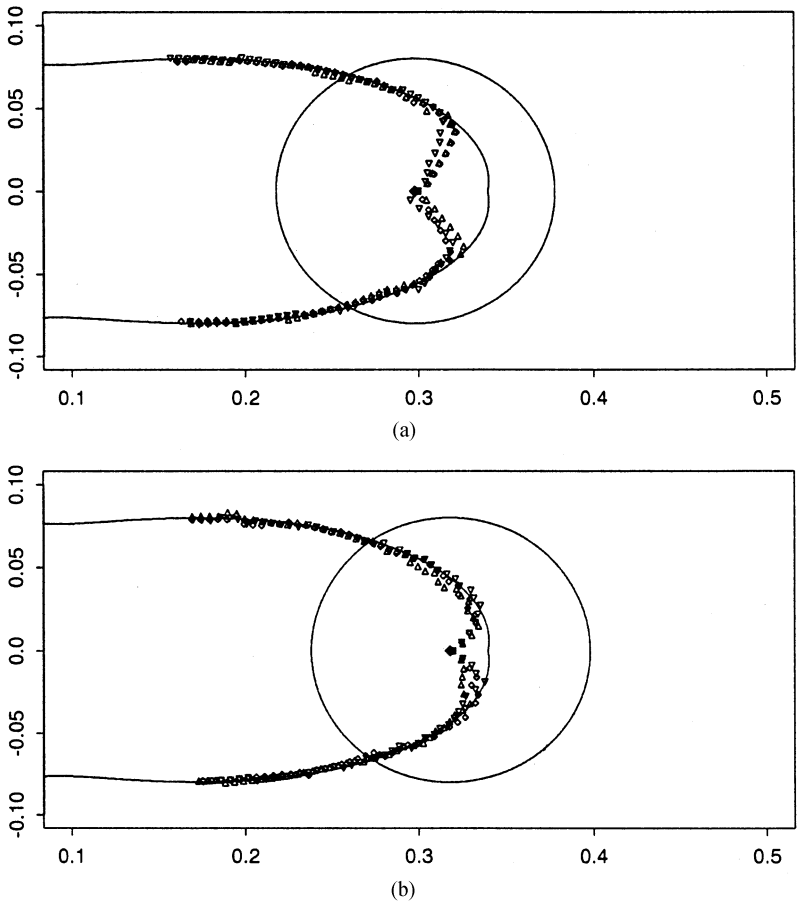

FIG. 3. *Panels show the effects of choice of* $\widehat{Q}$ *on behavior of the tracking estimate. Three realizations are depicted on each panel, represented, respectively, by empty diamonds, by empty triangles pointing upward and by empty triangles pointing downward. The starting point* $\widehat{Q}$ *is represented by a filled diamond. It moves steadily from the left-hand side to the right-hand side of the curve as we progress through panels* (a)–(d).

as (x_1, y_1) converges to (x, y) through values in $\mathscr{R}_R$, then

$$\inf_{(x, y)\in\mathscr{C}} |f_L(x, y) - f_R(x, y)| > 0.$$

Assume the errors ε_i are independent and identically distributed with zero mean and a distribution that has finite moment generating function in the neighborhood of the origin; call this condition (C_{err}). Suppose K is a nonnegative function supported on $[0, 1]$, not vanishing on $[0, 1)$ and Lipschitz continuous on $[0, \infty)$. Call these conditions (C_{ker}).

We assume the following condition, (C_{pp}), of the point process $\mathscr{P}_\nu$ from which come the points (X_i, Y_i): either $\mathscr{P}_\nu$ is homogeneous Poisson with inten-

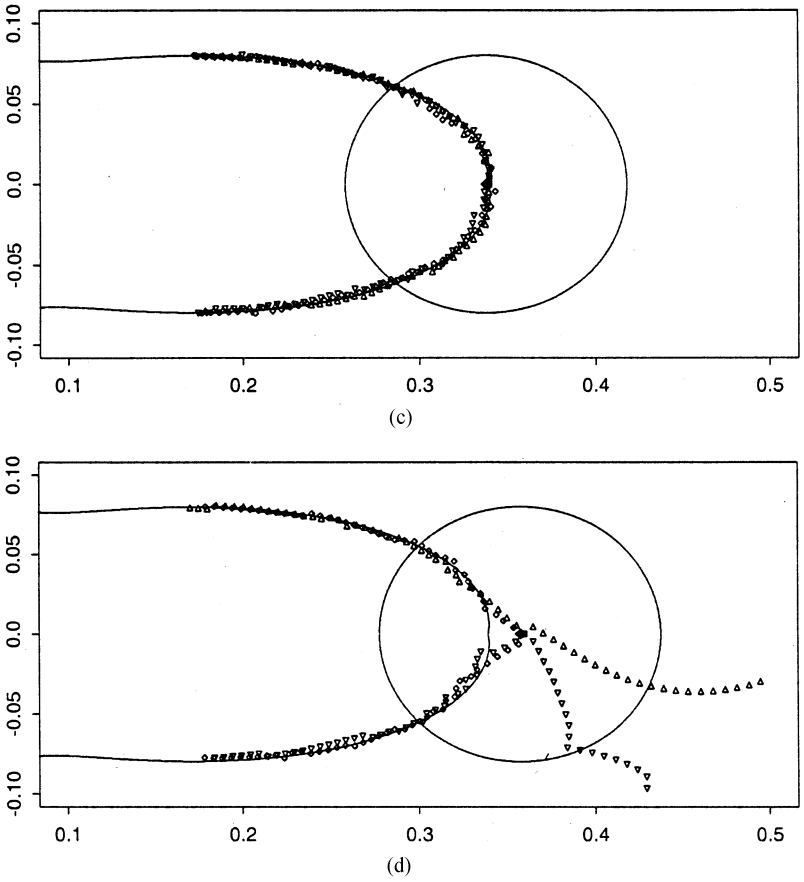

FIG. 3. *Continued*

sity ν, or it is supported on a regular triangular, square or hexagonal lattice with ν points per unit area; call this (C_{pp}). Of the bandwidth h we assume the following condition, (C_{bw}): $h = h(\nu) \to 0$ as $\nu \to \infty$; in the Poisson case, $(\log \nu)^2/(\nu h^3) \to 0$; and in the lattice case, $1/(\nu h^4) \to 0$. (A logarithmic factor is not required in the lattice case; see Section 4.1 for discussion.)

THEOREM. *Assume that conditions* (C_{bw}), (C_{err}), (C_{ker}), (C_{pp}) *and* (C_{rs}) *are satisfied, that* $\widehat{Q}$ *is within* $C_1 h^2$ *of* Q *and that the slope* $\hat{\theta}$ *is within* $C_2 h$ *of the slope of* $\mathscr{C}$ *at* Q *for some constants* $C_1, C_2 > 0$. *Then with probability* 1, *for some* $C_3 > 0$, $\widehat{\mathscr{C}}$ *is contained in the envelope of all points that lie within* $C_3 h^2$ *of* $\mathscr{C}$. *Moreover, with probability* 1 *the algorithm terminates after* $O(h^{-2})$ *steps*.

COROLLARY. *Assume the conditions of the theorem. Then, provided* $h \to 0$ *and either* $\nu^{-1/3}(\log \nu)^{2/3} = o(h)$ (*in the Poisson process case*) *or* $\nu^{-1/4} = o(h)$ (*for a regular lattice*), *the conclusion of the theorem holds. Therefore,*

any convergence rate that is slower than $\nu^{-2/3}(\log \nu)^{4/3}$ *or* $\nu^{-1/2}$, *respectively*, *is achievable almost surely by the estimator* $\widehat{\mathscr{C}}$.

REMARK 3.1 (Proximity to optimal convergence rates in the deterministic case). That the convergence rate $O(\nu^{-2/3})$, modulo logarithmic factors, is optimal in the Poisson case follows from work of Korostelev and Tsybakov [(1993), Section 5.3] and Mammen and Tsybakov (1995). The optimal rate for lattice data is $O(\nu^{-1/2})$.

Likewise, any convergence rate that is slower than $\nu^{-1/3}(\log \nu)^{2/3}$ or $\nu^{-1/4}$, in the Poisson and lattice cases respectively, is achievable in approximating the slope of $\mathscr{C}$ by the value of θ, computed by least-squares as at (2.1). These are within at most logarithmic factors of the best that are possible in a minimax sense. It is essentially this property which guarantees that, as claimed in the theorem, the algorithm concludes after $O(h^{-2})$ steps, since it ensures that the direction in which we move when estimating the curve by a sequence of points on an $O(h^2)$ grid is close to being the correct one, that is, along the curve.

REMARK 3.2 (Oversmoothing). Condition (C_{bw}) is deliberately constructed so as to produce enough oversmoothing to allow a relatively simple asymptotic description of $\widehat{\mathscr{C}}$. We may smooth a little less, producing a slightly faster convergence rate in the Poisson case (although only by a logarithmic factor), at the expense of a more complex asymptotic description of $\widehat{\mathscr{C}}$ and a longer proof.

REMARK 3.3 (More general point processes). The theorem may be generalized to include the cases where $\mathscr{P}_\nu$ is a Poisson cluster process or a jittered grid process. The latter were discussed by Korostelev and Tsybakov (1993). In all these settings the conditions on the bandwidth, and the conclusions of the theorem and the corollary, are the same as in the Poisson process case. For Poisson and Poisson cluster processes, the intensity need not be constant. It is sufficient that the process have intensity $\nu\,\lambda(x, y)$ at each point (x, y) in the plane, where λ is kept fixed as ν diverges, and is bounded away from zero and has bounded first derivatives in $\mathscr{R}$.

REMARK 3.4 (Estimating the starting point and slope). Assume $\mathscr{C}$ cuts the y axis $\mathscr{A}$, with equation $x = 0$, at a point Q where $\mathscr{C}$ is not tangential to the axis, and that Q is unique within $\mathscr{R} \cap \mathscr{A}$. Suppose too that $\mathscr{C}$ has two bounded derivatives in a neighborhood of Q. Let $h = h(\nu)$ denote a bandwidth sequence that satisfies (C_{bw}) in the Poisson case, that is, $(\log \nu)^2/(\nu h^3) \to 0$. Below we describe, in the Poisson case, a method for calculating an estimate $\widehat{Q}$ of Q which lies on $\mathscr{A}$ and, under the above conditions, achieves $O(h^2)$ accuracy with probability 1. It involves employing a univariate change-point method to compute a pilot estimator $\widehat{Q}_1$, with an error of order $h^{(3/2)-\varepsilon}$ for any given $\varepsilon > 0$, and then refining this to $\widehat{Q}$ by using the least-squares criterion at (2.1) with (x, y) there taken equal to the coordinates of $\widehat{Q}_1$. In addition, the approach we

describe below provides a starting orientation which achieves $O(h)$ accuracy with probability 1, so that the estimators obtained by this procedure satisfy the assumptions of the theorem, almost surely.

Let $h_1 = \nu^{-(1/2)+\Delta}$ where $0 < \Delta < \frac{1}{2}$, and project vertically onto $\mathscr{A}$ all those points $(X_i, Y_i) \in \mathscr{P}_\nu$ within the band $|X_i| \le h_1$. This produces a linear point process with intensity $2\nu h_1$. From these data and their associated Z_i's, a standard change-point estimator [see, e.g., Müller (1992), Eubank and Speckman (1994)] may be employed to derive an estimator $\widehat{Q}_1$ of Q, with coordinates $(0, \hat{y})$ say, subject to an error of $O\{(\nu h_1)^{-1+\delta} + h_1\}$ almost surely, for all $\delta > 0$. In view of our choice of h_1, this is of order $\nu^{-(1/2)+\Delta}$. Since h satisfies $(\mathrm{C}_{\mathrm{bw}})$ then the error equals $O(h^{(3/2)-\varepsilon})$ for each $\varepsilon > 3\Delta$.

Choose $\Delta \in (0, \frac{1}{6})$, or alternatively $\varepsilon \in (0, \frac{1}{2})$, and substitute the coordinates $(x, y) = (0, \hat{y})$ of $\widehat{Q}_1$ into the definition of $S(\xi, \eta, \theta)$ at (2.1). Choosing (ξ, η, θ) to minimize S produces the coordinate pair (ξ, η) of a point $\widehat{Q}_2$ which, with probability 1, is within $O(h^2)$ of Q, and a starting orientation θ which, with probability 1, is within $O(h)$ of the slope of $\mathscr{C}$ at Q. (Methods outlined in Section 4.2 may be used to derive these results.) Taking $\widehat{Q} = (0, \eta)$ and $\hat{\theta} = \theta$, we obtain the desired starting point and starting orientation for the algorithm.

4. Proof of the theorem.

4.1. *Intuitive outline of the proof.* First we consider the Poisson case. Let $\mathscr{D}$ be a disc of radius h in the plane, centered at (x, y) and with points (X_i, Y_i) of the Poisson process $\mathscr{P}_\nu$ distributed through it. The disc represents the region of support of the kernel function $K\{\|(x, y) - (\cdot, \cdot)\|/h\}$ appearing in (2.1). Recall that the line $\mathscr{U} = \mathscr{U}(\xi, \eta, \theta)$ represents the boundary where the jump in the local linear model fitted at (2.1) occurs.

Suppose the part of $\mathscr{C}$ which lies within $\mathscr{D}$ is distant $C_1 h^2$, at its furthest point, from that part of $\mathscr{U}$ that lies within the disc. (The symbols $C_1, C_2 \ldots$ here and below denote positive constants.) Then within the disc there is a region, with area $O(h^3)$, between the line $\mathscr{U}$ and the curve $\mathscr{C}$. To see why the area is of this size, note that the region can be approximated by a rectangle of which one side is of length $C_2 h$ (the order of the length of a diameter of the disc; the line $\mathscr{U}$ will not be far from being a diameter), and the other is of length $C_3 h^2$ (the order of the distance between $\mathscr{U}$ and $\mathscr{C}$).

The region can be thought of as representing errors arising from two sources: (a) approximating the curve $\mathscr{C}$ by a straight line in our local linear model and (b) putting the approximating line in the wrong position. When we fit the local linear model we can detect errors of types (a) and (b) if the number of Poisson-distributed points falling within the region is large enough. Since the region has area $O(h^3)$ then the number of points there is $O(\nu h^3)$, and so we can detect the errors if h is chosen so that νh^3 is sufficiently large.

We can determine what is required by "sufficiently large" by arguing as follows. For small h, the response surface will be locally constant on either

side of the jump, and we can estimate the two constants to within $o_p(1)$ by fitting the local linear model. Since the jump is $O(1)$ and not $o(1)$ then we do not have to be very accurate when estimating the constants. Now, the performance of the least-squares fit will deteriorate noticeably as soon as the number of points in the region increases beyond roughly $O(1)$, because then the fitted constant there will either be the one for the high side of the jump when it should be that for the low side, or vice versa. Taking account of the need to control moderate deviations, it turns out that "roughly $O(1)$" has to be interpreted as a power of $\log \nu$.

Hence, we conclude that νh^3 should be no smaller than a certain power of $\log \nu$; or equivalently, that the diameter h of the original disc centered at (x, y) should be no smaller than $\nu^{-1/3}$, multiplied by a logarithmic factor. As the argument above suggests, if h is chosen in this way, then the least-squares fitting procedure proposed at (2.1) can detect departures of up to $O(h^2)$ from the true curve. And, by its nature, having detected the departure the least-squares procedure applies an appropriate correction. It is critical to this argument that we start the curve-tracking procedure at a point which is within $O(h^c)$ of the true curve, for some $c > 1$. But once that is done, the fact that we can detect each time the curve estimate wanders beyond $C_5 h^2$ away from the true curve (for some $C_5 > 0$), and correct for it, means that we can stay within the range $O(h^2)$ on all subsequent steps.

Similarly it can be shown that, provided the grid of θ values contains at least $O(h^{-1})$ elements, the estimate of the slope of $\mathscr{C}$ which is given by the value of θ defined by minimizing $S(\xi, \eta, \theta)$ at (2.1), is within $O(h)$ of the true slope. [The $O(h)$ and $O(h^2)$ assertions here and above apply with probability 1, uniformly along the length of $\mathscr{C}$.]

When points of $\mathscr{P}_\nu$ are distributed on a lattice, rather than randomly distributed, an extra restriction is necessitated by the fact that regions which are narrower than a constant multiple of $\nu^{-1/2}$ (the distance between adjacent rows of points) might not receive any points at all. In particular, in order for the region whose width is h^2 to be guaranteed to have enough points to sustain our earlier argument, it is necessary that h^2 be an order of magnitude larger than $\nu^{-1/2}$. That is, in the case of lattice-distributed design points, h must be an order of magnitude larger than $\nu^{-1/4}$, as assumed in condition $(\mathrm{C}_{\mathrm{bw}})$.

This point is related to the reason why a $\log \nu$ factor is not required in the condition $(\mathrm{C}_{\mathrm{bw}})$ in the lattice case. Indeed, theoretical arguments in the lattice case demand only that the number of lattice points within a rectangle with dimensions $\delta_1 h \times \delta_2 h^2$ diverges to infinity at a rate faster than $(\log \nu)^2$, for any fixed $\delta_1, \delta_2 > 0$. Provided the edge length of the lattice is of strictly smaller order than h^2, or equivalently, provided $\nu h^4 \to \infty$, the number of points in the rectangle is asymptotic to $\delta_1 \delta_2 \nu h^3$, which in the lattice case is polynomially large in ν.

4.2. *Details of the proof.* Let (x, y) be a point on the grid $\mathscr{G}$ of potential estimates of points on $\mathscr{C}$, and let (ξ, η) be any point on $\mathscr{G}$ which is within $h^{3/2}$ of (x, y). (We could replace $h^{3/2}$ by h^c for any $1 < c < 2$.) Let $\mathscr{T} = \mathscr{T}(\xi, \eta)$

denote the disc of radius $2h$ centered at (ξ, η), and let $\mathscr{U}(\xi, \eta, \theta)$ be the line passing through (ξ, η) in the direction of the unit vector θ. Then $\mathscr{U}(\xi, \eta, \theta)$ divides $\mathscr{T}$ into two half-discs, $\mathscr{T}_1(\xi, \eta, \theta)$ and $\mathscr{T}_2(\xi, \eta, \theta)$, say. For $j = 1, 2$ let $\sum_i^{(j)}$ denote summation over indices i such that $(X_i, Y_i) \in \mathscr{T}_j(\xi, \eta, \theta)$. Write $\mathscr{G}_1$ for the set of all possible values of $(x, y, \xi, \eta, \theta)$. [For simplicity of notation we do not indicate dependence on (x, y), and until the paragraph containing (4.7) and (4.8) we also suppress dependence on (ξ, η, θ).] Put

$$K_i = K\{\|(x, y) - (X_i, Y_i)\|/h\}, \qquad \kappa^{(j)} = \sum_i^{(j)} K_i,$$

$$\mu^{(j)}\,\kappa^{(j)} = \sum_i^{(j)} f(X_i, Y_i)\,K_i, \qquad \bar{\varepsilon}^{(j)}\,\kappa^{(j)} = \sum_i^{(j)} \varepsilon_i\,K_i,$$

$$M^{(j)} = \big(\kappa^{(j)}\big)^{-1} \sum_i^{(j)} Z_i K_i = \mu^{(j)} + \bar{\varepsilon}^{(j)},$$

$$\delta_i^{(j)} = f(X_i, Y_i) - \mu^{(j)}, \qquad \Delta^{(j)} = \sum_i^{(j)} \delta_i^{(j)}\,\varepsilon_i\,K_i, \qquad S_1 = \sum_i^{(0)} \varepsilon_i^2\,K_i,$$

$$S_2 = \sum_{j=1,2} \sum_i^{(j)} \big(\delta_i^{(j)}\big)^2\,K_i, \qquad S_3 = \sum_{j=1,2} \big(\bar{\varepsilon}^{(j)}\big)^2\,\kappa^{(j)},$$

$$S_4 = \sum_{j=1,2} \Delta^{(j)}, \qquad S_5 = \sum_{j=1,2} \bar{\varepsilon}^{(j)} \sum_i^{(j)} \delta_i^{(j)}\,K_i.$$

In this notation, the quantity $S(\xi, \eta, \theta)$ defined at (2.1) is given by

$$S(\xi, \eta, \theta) = \sum_{j=1,2} \sum_i^{(j)} \big(Z_i - M^{(j)}\big)^2\,K_i = S_1 + S_2 - S_3 + 2\,S_4 - 2\,S_5. \tag{4.1}$$

[In passing from (2.1) to (4.1) we have found the values of a and b that minimize the sum of squares on the right-hand side of (2.1), and substituted them back into the formula.] By the Cauchy–Schwarz inequality, $|S_5| \le (S_2 S_3)^{1/2}$, and so

$$|S(\xi, \eta, \theta) - (S_1 + S_2)| \le S_3 + 2|S_4| + 2\,(S_2\,S_3)^{1/2}. \tag{4.2}$$

Next we develop bounds to $T_1 \equiv \bar{\varepsilon}^{(j)}\kappa^{(j)}$ and $T_2 \equiv \Delta^{(j)}$. Both T_1 and T_2 may be written in the form $T = \sum_i \varepsilon_i w_i$, where the weights w_i are such that $w = \sup_i |w_i|$ and $W^2 = \sum_i w_i^2$ are both finite. Since, by (C_{err}), the distribution of the errors ε_i has a finite moment generating function γ, say, in a neighborhood of the origin, then there exist constants $D_1, D_2 > 0$ such that $|\log \gamma(t)| \le D_1\,t^2$ whenever $|t| \le D_2$. Given $u > 0$, put $x = uW$ and $t = \min\{u/(2D_1W), D_2/w\}$. Then, by Markov's inequality, using condition (C_{err}) and noting that $|tw_i| \le D_2$ for each i, we have that $P'(T \ge$

$x) \le E'(e^{tT-tx}) \le \exp(D_1 t^2 W^2 - tx)$, where P' and E' denote probability and expectation conditional on the points of $\mathscr{P}_\nu$. Considering separately the cases $u \le 2D_1D_2W/w$ and $u > 2D_1D_2W/w$ we may show that for all $u > 0$, $D_1t^2W^2 - tx \le -C_1 u \min(u, W/w)$, where $C_1 > 0$ depends only on D_1 and D_2. Furthermore, in the cases $T = T_1$ and $T = T_2$ we have, respectively, $w \le \sup K$ and $w \le 2(\sup K)(\sup|f|)$. From these results, and their analogues in the opposite tail, we deduce that in either case,

$$P'(|T| > u\,W) \le 2\exp\{-C_2\,u\,\min(u, W)\}. \tag{4.3}$$

Note that in the case of T_k, $W = W_k$ where $W_1^2 = \sum_i^{(j)} K_i^2$ and $W_2^2 = \sum_i^{(j)} (\delta_i^{(j)})^2 K_i^2$. When ambiguity could otherwise occur we shall write T_k as $T_k^{(j)}$ and W_k as $W_k^{(j)}$. In this notation, and using (4.3), we have

$$\begin{aligned} &P'\{(S_2S_3)^{1/2} \ge (2S_2 \sup K)^{1/2}\,u\} \\ &\quad = P'\{S_3 \ge 2(\sup K)\,u^2\} \le \sum_{j=1,2} P'\{(T_1^{(j)})^2/\kappa^{(j)} > (\sup K)\,u^2\} \\ &\quad \le \sum_{j=1,2} P'(T_1^{(j)} \ge W_1^{(j)}\,u) \le 2\sum_{j=1,2} \exp\{-C_2\,u\,\min(u, W_1^{(j)})\}. \end{aligned}$$

Also, provided $u \ge 1$,

$$\begin{aligned} P'\{|S_4| > 2\,u^2 \max(S_2^{1/2}, 1)\} &\le \sum_{j=1,2} P'\{|T_2^{(j)}| > u^2 \max(W_2^{(j)}, 1)\} \\ &\le 2\sum_{j=1,2} \exp\Big[-C_2\,u^2 \max(1, 1/W_2^{(j)}) \\ &\qquad \times \min\{u^2 \max(1, 1/W_2^{(j)}), W_2^{(j)}\}\Big] \\ &\le 4\exp(-C_2\,u^2). \end{aligned}$$

Therefore, by (4.2), if $u \ge 1$,

$$|S(\xi, \eta, \theta) - (S_1 + S_2)| \le C_3\,u^2 \max(S_2^{1/2}, 1)\,Z_1, \tag{4.4}$$

where $C_3 > 0$ depends only on $\sup K$, and the nonnegative random variable $Z_1 = Z_1(u)$ satisfies

$$P'(Z_1 > 1) \le 6\sum_{j=1,2} \exp\{-C_2\,u\,\min(u, W_1^{(j)})\}. \tag{4.5}$$

If the points in $\mathscr{P}_\nu$ are Poisson-distributed then, for all $C > 0$,

$$P\Big[\inf_{(x, y, \xi, \eta, \theta) \in \mathscr{G}_1} \{\min(W_1^{(1)}, W_1^{(2)})\}^2 \ge C_4\,\nu\,h^2\Big] = 1 - O(\nu^{-C}), \tag{4.6}$$

where $C_4 > 0$ depends only on K. To appreciate why, use Markov's inequality to prove that $P\{|(W_1^{(j)})^2 - E(W_1^{(j)})^2| > \varepsilon\nu h^2\} = O(\nu^{-C})$ for all $\varepsilon, C > 0$, uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$. Direct calculation shows that $E(W_1^{(j)})^2/(\nu h^2)$ is

bounded away from 0 uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$. From these properties and the fact that $\mathscr{G}_1$ has $O(\nu^C)$ elements for some $C > 0$, we obtain (4.6).

When the points in $\mathscr{P}_\nu$ are on a triangular, square or hexagonal lattice, condition (C_{bw}) and the fact that the lengths of lattice edges are asymptotic to a constant multiple of $\nu^{-1/2}$ imply that for all sufficiently large ν, $(W_1^{(j)})^2/(\nu h^2)$ is bounded away from 0 uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$. In this case, (4.6) holds in a degenerate form, with its right-hand side replaced by 1, for some $C_4 > 0$.

From (4.4), (4.5) and (4.6), and the fact that $\nu h^2/\log \nu \to \infty$ under condition (C_{bw}), we may show that for any $C_5 > 0$ there exists $C_6 > 0$ such that, uniformly in choices of $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$,

$$\begin{aligned} &\bigl|S(\xi, \eta, \theta) - \{S_1 + S_2(\xi, \eta, \theta)\}\bigr| \\ &\qquad \le C_6 (\log \nu) \max\bigl\{S_2(\xi, \eta, \theta)^{1/2}, 1\bigr\} Z_2(\xi, \eta, \theta), \end{aligned} \tag{4.7}$$

$$P\{Z_2(\xi, \eta, \theta) > 1\} \le C_6 \nu^{-C_5}. \tag{4.8}$$

[Take u in (4.4) and (4.5) to equal a sufficiently large constant multiple of $(\log \nu)^{1/2}$.] From this point we explicitly state the dependence of S_k on (ξ, η, θ).

Since the number of elements of $\mathscr{G}_1$ is only polynomially large in ν, and since the constant C_5 at (4.8) may be taken arbitrarily large, then by (4.7), (4.8) and the Borel–Cantelli lemma we have that with probability 1, for all sufficiently large ν,

$$\begin{aligned} &|S(\xi, \eta, \theta) - \{S_1 + S_2(\xi, \eta, \theta)\}| \le C_7 (\log \nu) \max\bigl\{S_2(\xi, \eta, \theta)^{1/2}, 1\bigr\} \\ &\qquad\qquad \text{for all } (x, y, \xi, \eta, \theta) \in \mathscr{G}_1. \end{aligned} \tag{4.9}$$

Likewise, going back to the bound at (4.3) and taking u equal to a sufficiently large constant multiple of $(\log \nu)^{1/2}$, we may prove that with probability 1, for all sufficiently large ν,

$$\bigl|\bar{\varepsilon}^{(j)}(\xi, \eta, \theta)\bigr| \, \kappa^{(j)}(\xi, \eta, \theta)^{1/2} \le C_7 \log \nu \quad \text{for all } (x, y, \xi, \eta, \theta) \in \mathscr{G}_1. \tag{4.10}$$

Let D (depending only on B_1) denote an upper bound, uniformly in $(\xi, \eta) \in \mathscr{G}$ and in all directions, for the absolute value of the first directional derivative of f at (ξ, η). Write $\kappa^{(j,1)} = \kappa^{(j,1)}(\xi, \eta, \theta)$ for the sum of K_i over indices i such that $(X_i, Y_i) \in \mathscr{T}_j(\xi, \eta, \theta)$ is on the left-hand side of $\mathscr{C}$, and let $\kappa^{(j,2)}$ denote the same for the right-hand side of $\mathscr{C}$. Let $b = b(\xi, \eta)$ be the integral mean value, taken over $(u, v) \in \mathscr{C} \cap \mathscr{T}(\xi, \eta)$, of $|f_L(u, v) - f_R(u, v)|$, with $b = 0$ if $\mathscr{C} \cap \mathscr{T}(\xi, \eta)$ is empty. Then, if $(X_i, Y_i) \in \mathscr{T}(\xi, \eta)$ is on the left-hand side of $\mathscr{C}$,

$$b \, \kappa^{(j,2)} \bigl(\kappa^{(j)}\bigr)^{-1} - 4 \, D \, h \le \bigl|\delta_i^{(j)}\bigr| \le b \, \kappa^{(j,2)} \bigl(\kappa^{(j)}\bigr)^{-1} + 4 \, D \, h,$$

and if (X_i, Y_i) is on the right-hand side of $\mathscr{C}$ then the same is true provided $\kappa^{(j,2)}$ is replaced by $\kappa^{(j,1)}$. [Here and in (4.11) below we suppress dependence of $\kappa^{(j)}$, $\kappa^{(j,1)}$, $\kappa^{(j,2)}$ and $\delta_i^{(j)}$ on (ξ, η, θ).] Squaring, multiplying by K_i, adding over i such that $(X_i, Y_i) \in \mathscr{T}_j(\xi, \eta, \theta)$ and noting that

$$\bigl(\kappa^{(j,2)}/\kappa^{(j)}\bigr)^2 \kappa^{(j,1)} + \bigl(\kappa^{(j,1)}/\kappa^{(j)}\bigr)^2 \kappa^{(j,2)} = \kappa^{(j,1)} \, \kappa^{(j,2)}/\kappa^{(j)},$$

we deduce that

$$
\begin{aligned}
&b\,(b - 16\,Dh)\,\kappa^{(j,1)}\,\kappa^{(j,2)}\,\bigl(\kappa^{(j)}\bigr)^{-1} \\
&\qquad \le \sum_i^{(j)} \bigl(\delta_i^{(j)}\bigr)^2\,K_i \le b\,(b + 16\,Dh)\,\kappa^{(j,1)}\,\kappa^{(j,2)}\,\bigl(\kappa^{(j)}\bigr)^{-1} \\
&\qquad\qquad +16\,\kappa^{(j)}\,D^2\,h^2.
\end{aligned} \tag{4.11}
$$

Next we prove that for a constant $C_8 > 0$, and for $j = 1, 2$,

$$
\bigl(\kappa^{(j)} - C_8 \nu h^2\bigr)\,\bigl(\nu h^2\bigr)^{-1} = o(1), \tag{4.12}
$$

where the random variable represented by "$o(1)$" is of that size uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$, with probability 1. To derive (4.12), consider first the case where $\mathscr{P}_\nu$ is a Poisson process. Condition on the number $N^{(j)}$ of nonvanishing terms in the series defining $\kappa^{(j)}$, and then use Bernstein's inequality to bound the probability

$$
\pi\bigl(C, N^{(j)}\bigr) \equiv P\bigl\{\bigl|\kappa^{(j)} - E\bigl(\kappa^{(j)}\bigm|N^{(j)}\bigr)\bigr| > C\,\mathrm{var}\bigl(\kappa^{(j)}\bigm|N^{(j)}\bigr)^{1/2}\log\nu\bigm|N^{(j)}\bigr\}
$$

in the case $N^{(j)} > (\log\nu)^2$, to prove that if $C > 0$ is given then $C' > 0$ may be chosen so large that $E\{\pi(C', N^{(j)})\} = O(\nu^{-C})$. Also, $E(\kappa^{(j)}|N^{(j)})$ equals a constant multiple of $N^{(j)}$, and for each $C > 0$,

$$
P\bigl\{\bigl|N^{(j)} - E\bigl(N^{(j)}\bigr)\bigr| > C\,\bigl(\mathrm{var}N^{(j)}\bigr)^{1/2}\log\nu\bigr\} = O\bigl(\nu^{-C}\bigr).
$$

Note too that $\mathrm{var}(\kappa^{(j)}|N^{(j)})$ and $\mathrm{var}(N^{(j)})$ are bounded above and below by constant multiples of $N^{(j)}$ and νh^2, respectively, and that $E(N^{(j)}) = \frac{1}{2}\pi\nu h^2$. (The variable $N^{(j)}$ is Poisson-distributed with this mean.) Combining these results with the Borel–Cantelli lemma, and exploiting the fact that the number of elements of $\mathscr{G}_1$ is only polynomially large in ν, we obtain (4.12) in the Poisson case.

The context where $\mathscr{P}_\nu$ is a lattice may be treated more directly, since there $\kappa^{(j)}$ is a deterministic sum. It is asymptotic to a constant multiple of νh^2, uniformly in (x, y, ξ, η). In the lattice case the assertion "with probability 1" is not required when describing (4.12).

Let $\mathscr{I} = \mathscr{I}(\xi, \eta, \theta)$ be the line segment of length $4h$ defined as the diameter of $\mathscr{T} = \mathscr{T}(\xi, \eta)$ that is aligned in direction θ, and put $\mathscr{C}_\mathscr{T} = \mathscr{C} \cap \mathscr{T}$. Let $R^{(j,k)}$ denote the portion of $\mathscr{T}_j(\xi, \eta, \theta)$ that is on the left-hand side of $\mathscr{C}$ when $k = 1$, and on the right-hand side when $k = 2$. Assume for the time being that $\mathscr{C}_\mathscr{T}$ is perpendicularly distant no more than $h^{3/2}$ from $\mathscr{I}$ at the furthest point; we call this the "distance" assumption. When it holds, one of $R^{(j,1)}$ and $R^{(j,2)}$ is of area not exceeding $4h^{5/2}$; let it be $R^{(j,k_j)}$. The other region is of area at least $2\pi h^2 - 4h^{5/2}$. [Recall that $\mathscr{T}(\xi, \eta)$ is of radius $2h$.] Let this region be $R^{(j,l_j)}$. The argument used to derive (4.12) may be employed to show that for $j = 1, 2$,

$$
\kappa^{(j,k_j)}\bigl(\nu h^2\bigr)^{-1} = o(1) \tag{4.13}
$$

uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$ for which the "distance" assumption holds, with probability 1. Combining this result with (4.12) we conclude that $\kappa^{(j, l_j)}/\kappa^{(j)} = 1 + o(1)$, where the "$o(1)$" term has the same interpretation as that at (4.13). Hence,

$$\sum_{j=1, 2} \kappa^{(j, 1)} \kappa^{(j, 2)} \bigl(\kappa^{(j)}\bigr)^{-1} = \{1 + o(1)\} \bigl(\kappa^{(1, k_1)} + \kappa^{(2, k_2)}\bigr), \tag{4.14}$$

again with the same interpretation of "$o(1)$."

Let $\mathscr{J} = \mathscr{J}(\xi, \eta, \theta) = R^{(1, k_1)} \cup R^{(2, k_2)}$ denote the closed region within $\mathscr{T}(\xi, \eta)$ that is bordered by $\mathscr{C}_{\mathscr{T}}$, $\mathscr{I}$ and the perimeter of $\mathscr{T}(\xi, \eta)$. (Thus, $\mathscr{J}$ has a vertex at any point where $\mathscr{C}_{\mathscr{T}}$ intersects $\mathscr{I}$.) In the case where $\mathscr{P}_\nu$ is Poisson, let $A = A(\xi, \eta, \theta)$ equal the integral of $K\{\|(x, y) - (u, v)\|/h\}$ over $(u, v) \in \mathscr{J}$, and in the case where $\mathscr{P}_\nu$ is a lattice, let A be the average of this kernel over all grid points (u, v) that fall within $\mathscr{J}$. We claim that, with probability 1, uniformly in values $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$ for which the "distance" assumption holds,

$$\kappa^{(1, k_1)} + \kappa^{(2, k_2)} = \{1 + o(1)\} \nu A(\xi, \eta, \theta) + O\bigl\{(\log \nu)^2\bigr\}, \tag{4.15}$$

whence it follows from (4.14) that

$$\sum_{j=1, 2} \kappa^{(j, 1)} \kappa^{(j, 2)} \bigl(\kappa^{(j)}\bigr)^{-1} = \{1 + o(1)\} \nu A(\xi, \eta, \theta) + O\bigl\{(\log \nu)^2\bigr\}, \tag{4.16}$$

with probability 1, uniformly in the same range of values of x, y, ξ, η, θ.

To appreciate why (4.15) holds, consider first the case where $\mathscr{P}_\nu$ is Poisson. Let $A_1 = A_1(\xi, \eta, \theta)$ equal the area of the intersection, with the region $\mathscr{J}(\xi, \eta, \theta)$, of the unit disc centered at (x, y). Using the argument leading to (4.12) and (4.13) we may prove that, given $C > 0$, we can choose $C' > 0$ so large that for $j = 1, 2$,

$$P\bigl\{\bigl|\kappa^{(j, k_j)} - E\bigl(\kappa^{(j, k_j)}\bigr)\bigr| > C' (\nu A_1)^{1/2} \log \nu\bigr\} = O\bigl(\nu^{-C}\bigr)$$

uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$ for which the "distance" assumption holds and $\nu A_1 \geq (\log \nu)^2$. To treat the case where $\nu A_1 < (\log \nu)^2$ we note that when this inequality holds, the set of which A_1 is the area is contained within a set of which the area is $A_2 = \nu^{-1}(\log \nu)^2$. Hence, given $C > 0$, we may choose $C' > 0$ so large that

$$P\bigl[\bigl|\kappa^{(j, k_j)} - E\bigl(\kappa^{(j, k_j)}\bigr)\bigr| > C' \bigl\{(\nu A_1)^{1/2} + \log \nu\bigr\} \log \nu\bigr] = O\bigl(\nu^{-C}\bigr)$$

uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$ for which the "distance" assumption holds, this time without regard for the sign of $\nu A_1 - (\log \nu)^2$. Result (4.15) follows from this property and the fact that $E(\kappa^{(1, k_1)} + \kappa^{(2, k_2)})$ equals νA. The case where $\mathscr{P}_\nu$ is defined on a lattice may be treated by counting. In this context,

$$E\bigl(\kappa^{(1, k_1)} + \kappa^{(2, k_2)}\bigr) = \nu A + o(\nu A) + O\bigl\{(\log \nu)^2\bigr\}.$$

From (4.11), (4.12), (4.16) and the fact that $(\log \nu)^2/\nu h^3 \to 0$ we see that with probability 1, uniformly in values $(x, y, \xi, \eta, \theta) \in \mathscr{G}_1$ for which the "distance"

assumption holds,

$$S_2(\xi, \eta, \theta) = \sum_{j=1,2} \sum_{i}^{(j)} \left(\delta_i^{(j)}\right)^2 K_i \tag{4.17}$$
$$= \{1 + o(1)\}\, \nu\, b(\xi, \eta)^2\, A(\xi, \eta, \theta) + o\left(\nu h^3\right).$$

From (4.9) and (4.17) we conclude that with probability 1,

$$S(\xi, \eta, \theta) = S_1 + \{1 + o(1)\}\, \nu\, b(\xi, \eta)^2\, A(\xi, \eta, \theta) + o\left(\nu h^3\right), \tag{4.18}$$

uniformly in $(x, y, \xi, \eta, \theta) \in \mathscr{S}_1$ satisfying the "distance" assumption.

Note that while S_1 depends on (x, y) it does not depend on (ξ, η, θ). Therefore, (4.18) implies that if we minimize $A(\xi, \eta, \theta)$ rather than $S(\xi, \eta, \theta)$, with respect to (ξ, η, θ), we merely neglect terms in the target function which are of order $o(A + h^3)$, uniformly in $(x, y, \xi, \eta, \theta)$. Although these functions will generally have multiple minimizers, the following arguments are valid for any two respective choices, as assumption (C_{pp}) ensures that all minimizers are contained within a rectangle of size $O(h^2) \times O(h)$, with probability 1. For the purposes of the next paragraph we view the choice of $(\xi_0, \eta_0, \theta_0) \in \{\text{argmin}\, A(\xi, \eta, \theta)\}$ as a function of h, and proceed to show that the choice of (ξ_0, η_0) and θ_0 differs from the "correct" choices only by terms of order $O(h^2)$ and $O(h)$, respectively.

We start with noticing that assumption (C_{rs}) implies that $\mathscr{C}$ is locally quadratic. For greater ease of exposition we first consider the case when the area is not given kernel weights, although we continue to use the notation $A(\xi, \eta, \theta)$ for this modified target function. Geometric arguments may then be employed to show that if (ξ, η) is distant further than $O(h^2)$ from $\mathscr{C}$, with θ arbitrary, then $A(\xi, \eta, \theta)$ is strictly larger than $A(\xi_0, \eta_0, \theta_0)$, for sufficiently small h. Similarly it can be shown that if the sequence (ξ, η) is within $O(h^2)$ of the true curve but the slope component θ is such that there exists $C > 0$ with $\|\theta - \theta_0\| \geq Ch$, then for h small enough there exists a grid point (ξ', η') between the line and the parabola with the property $A(\xi', \eta', \theta_0) < A(\xi, \eta, \theta)$. (All foregoing statements hold with probability 1.) The previous arguments may be transferred to the case of kernel-weighting, if we too observe that as K is assumed to be Lipschitz continuous, the function $K\{\|(x, y) - (\cdot, \cdot)\|/h\}$ converges uniformly to $K\{\|(\xi, \eta) - (\cdot, \cdot)\|/h\}$ as $(x, y) \to (\xi, \eta)$. In summary, we thus have established that up to terms of sufficiently small order, the algorithm minimizes $A(\xi, \eta, \theta)$ with probability 1, and we claim that the theorem follows from this property.

To explain why, we note the following four properties. In points (a), (b) below we refer to the location component of the minimizer of $A(\xi, \eta, \theta)$, while (a′), (b′) refer to its slope. All of these four properties may be derived by somewhat cumbersome, although not exceedingly complicated, geometric considerations in the spirit of the previous paragraph. (a) There exist constants $0 < C_1 < C_2$ such that, if the point P with coordinates (x, y) at which we are situated at a given step is further than $C_2 h^2$ from $\mathscr{C}$, then the grid point (ξ, η) that results

from minimizing $A(\xi, \eta, \theta)$ is less than $C_1 h^2$ from $\mathscr{C}$. (b) If $C_3 > 0$ is given then there exists $C_4 = C_4(C_3) > 0$ such that, provided (x, y) is distant no more than $C_3 h^2$ from $\mathscr{C}$, the value of (ξ, η) that results from minimizing $A(\xi, \eta, \theta)$ is distant no more than $C_4 h^2$ from $\mathscr{C}$. In (a) and (b), distance of a point to $\mathscr{C}$ denotes shortest distance. The slope component θ which minimizes $A(\xi, \eta, \theta)$, has the following properties (a′), (b′) which, respectively, correspond to (a), (b). (a′) There exist constants $0 < C_5 < C_6$ such that, if the slope estimate ω from the previous step is further than $C_5 h$ from the true slope of the curve at the point of $\mathscr{C}$ that is closest to (ξ, η), then the grid point θ that results from minimizing $A(\xi, \eta, \theta)$ is less than $C_6 h$ from that slope. Here and in in (b′), we represent the true slope of the point on $\mathscr{C}$ to which we refer, by a unit vector such that the dot product with θ is nonnegative. (b′) If $C_7 > 0$ is given then there exists $C_8 = C_8(C_7) > 0$ such that, provided ω is distant no more than $C_7 h$ from the true slope on the corresponding point on $\mathscr{C}$ as given in (a′), the value of θ that results from minimizing $A(\xi, \eta, \theta)$ is distant no more than $C_8 h$ from that slope. [The constants $C_1, C_2, C_4(C_3), C_5, C_6, C_8(C_7)$ depend on the maximum curvature of $\mathscr{C}$ and on the constants B_1, B_2 in the definition of the grid of points (ξ, η) and θ, but for h sufficiently small they do not depend on h.]

Properties (a), (a′) and result (4.18) imply that (c) if the grid vertex (x, y) is further than $C_2 h^2$ from $\mathscr{C}$ then at the next step we move to a grid vertex (ξ, η) that is less than $C_1 h^2 + o(h^2)$ from $\mathscr{C}$. Properties (b), (b′) imply that (d) if at the n'th step we are at a point (x, y) which is distant no more than $C_3 h^2$ from $\mathscr{C}$, then (ξ, η) will be distant no more than $C_4(C_3) h^2 + o(h^2)$ from $\mathscr{C}$. Furthermore, in view of the uniformity of the remainder at (4.18), the $o(h^2)$ remainders here are uniformly small. Now, $\widehat{Q}$ and Q, and the corresponding slopes, are distant $O(h^2)$ and $O(h)$ apart, respectively. (The last assertion would have to be qualified "with probability 1" if Q and the slope at Q were estimated by a procedure such as described in Remark 3.4, but the conclusions, being subject to the same restriction, would not be affected thereby.) These initial conditions, and properties (c) and (d), imply that at each step the point estimate must uniformly lie within $O(h^2)$ of the curve. Also, due to the same initial conditions, the initial triplet (ξ, η, θ) (be it deterministic, or the result of an estimation procedure such as described in Remark 3.4) satisfies the "distance" assumption with probability 1. Then the fact that subsequent points are uniformly distant $O(h^2)$ from $\mathscr{C}$ implies that this is valid throughout, with probability 1.

Similarly, using (a′) and (b′) from two paragraphs earlier, it may be proved that the estimate of the slope of $\mathscr{C}$, provided by the value of θ obtained from minimizing with respect to (ξ, η, θ), is uniformly within $O(h)$ of the true slope of the curve at the point of $\mathscr{C}$ that is closest to (ξ, η). Therefore, the sign convention for passing from a slope estimate ω (a unit vector) at (x, y) to the slope estimate θ at (ξ, η) [see the paragraph containing (2.1)] ensures that successive point estimates progress steadily along the curve, on the grid $\mathscr{G}$

with edge width equal to a constant multiple of h^2, so that only $O(h^{-2})$ steps are required until termination within h of the opposite side of $\mathscr{R}$.

Acknowledgment. We are grateful to two reviewers for helpful comments on two earlier versions of the paper.

REFERENCES

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, 2nd ed. Wiley, New York.

EUBANK, R. L. and SPECKMAN, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-Point Problems* 130–144. IMS, Hayward, CA.

GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228.

HALL, P. and RAIMONDO, M. (1997a). Approximating a line thrown at random onto a grid. *Ann. Appl. Probab.* **7** 648–665.

HALL, P. and RAIMONDO, M. (1997b). Measuring the performance of boundary-estimation methods. In L_1-*Statistical Procedures and Related Topics.* 1–14. IMS, Hayward, CA.

HALL, P. and RAIMONDO, M. (1998). On global performance of approximations to smooth curves using gridded data. *Ann. Statist.* **26** 2206–2217.

HUERTAS, A. and MEDIONI, G. (1986). Detection of intensity changes with subpixel accuracy using Laplacian–Gaussian masks. *IEEE Trans. Patt. Anal. Mach. Intell.* **8** 651–664.

KNEIP, A., PARK, B. U. and SIMAR, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* **14** 783–793.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statist.* **82**. Springer, Berlin.

MARR, D. and HILDRETH, E. (1980). Theory of edge detection. *Proc. Roy. Soc. London Ser. B* **207** 187–217.

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.

MÜLLER, H.-G. and SONG, K. S. (1994). Maximin estimation of multidimensional boundaries. *J. Multivariate Anal.* **50** 265–281.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge Univ. Press.

QIU, P. (1998). Discontinuous surfaces fitting. *Ann. Statist.* **26** 2218–2245.

QIU, P. and YANDELL, B. (1997). Jump detection in regression surfaces. *J. Comput. Graph. Statist.* **6** 332–354.

RUDEMO, M. and STRYHN, H. (1994). Approximating the distribution of maximum likelihood contour estimators in two-region images. *Scand. J. Statist.* **21** 41–55.

SEIFORD, L. M. (1996). Data envelopment analysis: the evolution of the state-of-the-art, 1978–1995. *J. Productivity Anal.* **7** 99–138.

TITTERINGTON, D. M. (1985a). Common structure of smoothing techniques in statistics. *Internat. Statist. Rev.* **53** 141–170.

TITTERINGTON, D. M. (1985b). General structure of regularization procedures in image reconstruction. *Astronom. Astrophys.* **144** 381–387.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA
E-MAIL: christian.rau@anu.edu.au